DOCUMENT RESUME

ED 336 932 FL 019 353

AUTHOR Valdman, Albert
TITLE The Symposium on the Evaluation of Foreign Language Proficiency: Challenges to the Profession.
PUB DATE 88
NOTE 15p.; In: New Challenges and Opportunities. Dimension: Languages 87. Report of the Southern Conference on Language Teaching, p1-14, 1988. For the proceedings, see FL 019 352.
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Educational Trends; *Evaluation Criteria; *Language Proficiency; *Language Tests; *National Norms; Second Language Instruction; Student Evaluation; Testing
IDENTIFIERS ACTFL ETS Oral Proficiency Interview

ABSTRACT

This paper presents a report of a 1987 symposium on the Evaluation of Foreign Language Proficiency held in Bloomington, Indiana. Although much has been accomplished in language testing, much remains to be done before the language teaching profession has at its disposal a common means of measuring proficiency in the functional use of language in its various modalities that meets generally recognized standards of face validity, reliability, and fairness and that is, at the same time, relatively easy and economical to administer on a large scale. Even more remains to be done before the profession can substantiate any claims about direct links between the attainment of various proficiency levels and inventories of discrete linguistic, discourse, or sociolinguistic features that need to be mastered. Collaboration is needed in three areas: (1) development of operational definitions of basic terms; (2) proactive research for validation of the Oral Proficiency Interview (OPI) and American Council on the Teaching of Foreign Languages (ACTFL) proficiency guidelines; and (3) wider sharing of information, both domestically and internationally. Concerns about the direction taken by development of the OPI include overemphasis on evaluation of communicative interaction, risks in developing the OPI to measure low-level proficiency, and potential backlash against oral skill development. Valid and reliable tests for languages for specific purposes are needed. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Valdman, Albert.

1 TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED336932

The Symposium on the Evaluation of Foreign Language Proficiency:

Challenges to the Profession

Albert Valdman
Indiana University

Introduction

The ETS-ACTFL Oral Proficiency Interview (OPI) has had a profound impact on the foreign language (FL) teaching profession over the last four or five years.[1] There is scarcely any area of the field that has not been affected by this attempt to institute a national metric based on demonstrated proficiency in a FL and, more importantly, to define achievement in language instruction in terms of functional use rather than command of a specific body of material or, as Barbara Freed has put it, mere "seat-time."

But what has come to be called the "Proficiency Movement" involves much more than the development and dissemination of a particular instrument for the evaluation of demonstrable functional use of a FL. At the heart of the Movement are the ACTFL Guidelines which are viewed by some of its proponents as a veritable blueprint for the planning and delivery of FL instruction at the high school and university level. The generic guidelines define global proficiency levels in terms of three sets of judgments: (1) the ability to carry out certain functions using the target language, (2) the ability to deal with certain topic or content units, (3) the ability to produce a range of structural and discourse features with accuracy. Proponents of the Proficiency Movement stress the experiential rather than hypothetical nature of the Guidelines (Omaggio, p. 35), and they are confident that they can be used to specify inventories of elements whose control will lead students to attain specified levels of proficiency.

FL019353

BEST COPY AVAILABLE

> Because these guidelines and definitions represent actual rather than hypothetical language production, teachers can amend their expectations for students' linguistic and communicative development to conform to reality. The descriptions should prove useful in designing language programs, in the sense that they outline general performance criteria and specify the kinds of contexts and situations handled with reasonable ease at each level of competence. . . . Once we have identified, at least in terms of general performance criteria, what students have to know in order to function at or beyond a given level of competence, we will have come a long way toward knowing what it is we have to teach (Omaggio, p. 35).

There is general agreement that the Proficiency Movement launched by ACTFL has invigorated FL instruction and that is has had a beneficial backwash effect on a variety of areas from teacher training to material development and pedagogical procedures. Nonetheless several aspects of the OPI and the Guidelines have elicited severe reservations on the part of FL teaching methodologists and applied linguists who, like the proponents of the Proficiency Movement, place a high priority on the acquisition of functional skills by FL students. Unfortunately there has been little direct interaction between proponents of the Proficiency Movement and its critics. Yet this interaction could be mutually profitable, for it would lead to urgently needed research that would provide a firmer theoretical and empirical basis for not only a national FL proficiency metric but also for attempts to link levels of communicative ability in a FL to stages of acquisitional development. It was precisely in order to initiate this type of interaction that a Symposium on the Evaluation of FL Proficiency was held on the Bloomington Campus of Indiana University, March 4-6, 1987. To lay the groundwork for the Symposium, a workshop was organized in conjunction with the 1986 annual meeting of ACTFL in Dallas.

Based on the favorable reactions that followed the Workshop, approximately 30 American specialists of FL teaching methodology or applied linguistics were invited to prepare reports to be distributed in advance or to serve as discussants and resource persons at the Symposium.[2] They were selected from experts who have in various ways been concerned with the development of communicative ability in a FL and the assessment of that ability: (1) those involved in the development and the administration of the OPI and the elaboration of the ACTFL Guidelines; (2) those involved in the

training and certification of OPI raters; (3) those involved in the elaboration of communicatively-oriented curricula, syllabuses, teaching procedures, and materials; (4) those involved in the observation and analysis of second language learning. The participants represented a broad range of second and foreign language teaching situations: the commonly taught languages (French, German, and Spanish) and the major less commonly taught languages (Arabic, Chinese, Japanese, Indonesian, Russian, and Subsahara African languages) in the United States; indigenous languages of the Americas (including creoles); the less commonly taught languages of Europe; and the teaching of English as a second language. In addition to the participants several overseas observers were invited, especially from Western Europe where FL teaching specialists have taken the lead in developing communicatively-oriented programs and international schemes for the definition and assessment of proficiency levels, for example the Threshold Level and the Niveau Seuil. Invitations were also extended to a small number of representatives from professional groups and federal agencies involved in research and development in FL learning and in FL teaching.

The proceedings of the Symposium, containing all the position papers, discussant reports, and summaries of discussions are now available.[2] In this report, I will attempt to focus on the reservations expressed about the OPI and the Guidelines. I have selected four problem areas that were discussed at the Symposium: (1) the global nature of the OPI and related issues of validity, reliability, and fairness; (2) the view of communicative interaction inherent in the Guidelines; (3) the relationship between the OPI scales and the Guidelines and the nature of FL acquisition; (4) practical problems posed by the generalized used of the OPI. In the concluding section of this report I will mention some research directions identified at the Symposium.

Issues

1. The global nature of the OPI

For Lyle Bachman the central question in evaluating the OPI and the ACTFL-Interagency Language Roundtable (ILR) guidelines is whether they constitute a valid measure of communicative language proficiency, defined as the ability to conduct culturally embedded verbal interaction with speakers of the target language community. He answered in the negative because he believes the OPI confounds the competence to be measured with the method of measurement. Developers of the OPI are confronted with a

dilemma. They can aim for greater standardization and trade less content validity for more reliability of scoring; or they can expand the range of sample tasks and trade closer approximation to real linguistic interaction for reduced reliability.

Bachman views communicative language ability not as a holistic construct but one composed of three sets of constituent capabilities, each of which should be evaluated separately and each of which makes different predictions about an examinee's ability to perform specific communicative tasks. The Functional Trisection, which developers of the OPI view as its keystone, becomes for Bachman its chief liability, because the very use of the same performance sample to simultaneously evaluate content, function, and structural accuracy confuses test design with evaluation, content, and method. He also underscores the limited view of the nature of communicative proficiency reflected by evaluation guidelines that pay insufficient attention to pragmatic and sociolinguistic considerations. Instead he proposes a tripartite model in which linguistic structure, narrowly defined to encompass phonology, morphosyntax, and lexicon, forms only one of the four subcomponents of langauge competence. Bachman's scheme resembles the widely used Canale-Swain model in its stress on the rhetorical, pragmatic, and sociolinguistic features of interactive verbal communication. It differs, however, in that it isolates as a separate component control of motor skills, grouped under psychophysiological competence, and distinguishes more clearly between rules of language use, subsumed under pragmatic competence, and strategic competence. The latter identifies the ability to plan, execute, and assess socially embedded speech acts as a distinct and important part of communicative language ability.

The componential approach also characterizes the proficiency evaluation model devised specifically for Arabic and African langauges by Patrick Bennett and Ann Bierstecker. As is the case with Bachman's proposal, this scheme places greater emphasis on pragmatic and sociocultural aspects of verbal interactions than does the OPI.

Looking at the OPI from a narrower psychometric perspective Elana Shohamy raised questions about its validity and reliability. She granted that with extensive and **continuous** (underscoring mine) training of raters, acceptable levels of interrater reliability may be attained. However, she listed other types of reliability to which developers of the OPI have not paid sufficient attention: test-retest reliability, parallel forms reliability, and interinterviewer reliability. Because the interviewer is an integral part of the test such factors as interviewer attitude, personality, and rapport with the individual

being tested greatly influence results. Shohamy also stressed the narrow nature of the range of interactional competence tapped by the OPI, and she reported that a recent study she and her collaborators conducted showed only moderate correlations between four different types of verbal interactions: an interview, a role play, a reporting task, and a discussion. Observations such as these have led the Israel Ministry of Education to abandon an oral interview used for nearly twenty years in favor of a more varied scheme that samples verbal performance in the aforementioned four types of interaction. In addition, each part of the test is conducted by a different person.

Shohamy also mentions the need for applying criteria other than validity and reliability: utility, accuracy, feasibility, and fairness. For example, she points out that an important factor in the implementation of the OPI test is its positive backwash effect on instruction, particularly demands for higher levels of oral proficiency to be demonstrated by teachers and greater stress on the development of oral skills. With respect to the criterion of fairness, which Shohamy defines as insuring that a test be conducted legally, ethically, and with regard to the welfare of the individuals tested as well as others affected by test results, failure to sufficiently take it into account holds great potential risks in our litigious society.

The potential legal problems raised by the use of the OPI for purposes of establishing degree requirements and for teacher certification is precisely the issue addressed by Sally Magnan in her contribution. Referring to recent court cases, she stressed the need to establish the content and instructional validity of the test and its efficacy as a predictor of successful use of the target language and teaching competence. The link between proficiency in the language taught and pedagogical effectiveness will prove particularly difficult to establish in view of the multiplicity of objectives of general academic FL instruction and the vast array of knowledge and skills it demands on the part of teachers. Magnan cited the disquieting juridical precedent created by the 1979 case (Armstead versus Starkville Municipal Separate School District) in which the Graduate Record Examination was found unconstitutional as a criterion for the hiring and retention of high school teachers because its constituent tasks could not be proven to be directly related to good teaching. For these reasons, Magnan recommends that the use of the OPI or similar instruments for FL teacher certification should not be administered by degree granting institutions or employers of teachers but should be entrusted instead to professional teacher organizations.

John Clark and Ray Clifford outlined a program of research that would validate and provide a firmer empirical base for the OPI and the

ACTFL-ILR guidelines. They admit that, paradoxically, the OPI falls considerably short as a face- and content-valid measure of communicative interaction. They suggest research involving the correlation of OPI scores with a wider variety of more face- and content-valid activities and with highly specific tasks in employment and language use settings. Clark and Clifford also granted that the developers of the OPI must resolve problems on intra- and interrater reliability, and they pointed out that another area that deserves research is that of the relative resistance of communicative ability to attrition.

2. The nature of communicative interaction

Even in their new revised form, the Guidelines stress structural accuracy and functional adequacy to the detriment of sociolinguistic and discourse competence. In response to Heidi Byrnes' laudable attempt to incorporate these competences in the OPI scales, Claire Kramsch identified three weaknesses that make the OPI a less than adequate measure of what she termed interactional competence and that doom to reductionism all attempts to broaden the scope of rater judgments. First, because of the interactional nature of communication, the rater, as a cooperative partner in the elicitation of ratable samples, introduces an uncontrollable variable. As Bachman pointed out this results in a serious confounding of test content and test method. Second, basing the Guidelines on specific norms is incompatible with the evaluation of sociolinguistic competence which requires constant shifting of norms in reaction to the subtle changes in the social situation that inevitably occur during communicative interaction. Third, Kramsch pointed out the implicit sociocultural bias of a test that places a high premium on conversational management skills and decontextualized linguistic features associated with certain cultural levels and social groups. To eliminate these shortcomings, she suggested replacing the discrete behavioristic objectives, characteristic of traditional practice, with more abstract sociocultural constructs.

For Dale Lange one of the most serious problems for the implementation of the Guidelines is that level definitions related to the socio-cultural context are relegated to the upper end of the OPI scale. As Claire Kramsch (1986) has repeatedly stressed, genuine interactional competence, at any level, presupposes sociocultural knowledge.

> Without it, students are at the mercy of content structure, of the words they hear or read, without being able to recognize either the personal intentions or the cultural assumptions behind the words (p. 368).

The adaptation of the OPI and the Guidelines to the assessment of proficiency in languages other than the commonly taught Western European languages (one of the topics discussed at the Symposium) has heightened our awareness of how cultural considerations powerfully determine the degree of acceptable sociolinguistic deviance in learners' interlanguage. Unlike the Western European languages in general, those of the Far East, for example, require speakers to constantly situate themselves socially and choose appropriate linguistic variants with respect to the status of their interlocutors. It is clear that for such languages suitable guidelines must incorporate statements about sociolinguistic competence at the lowest proficiency levels.

3. The relationship between the OPI scales and the Guidelines and the nature of FL learning

As Nina Garrett pointed out in her position paper on the role of grammar in the Proficiency Movement:

> The Guidelines as rating level descriptions were originally developed on the basis of experiential data, reflecting the job-related requirements of the population being tested, rather than on the basis of adult classroom second language acquisition theory or theoretically motivated empirical research.

She delineated two mutually exclusive positions concerning the place of grammar in the OPI and the Guidelines: (1) relative levels of grammatical accuracy and sequences for the use of specific grammatical features are determined on the basis of experiential evidence, in which case no independent validation (for example, on the basis of psycholinguistic evidence) is necessary; or (2) such independent validation is used to legitimize claims about the developmental order of grammatical features and specification of levels of accuracy. Like several other Symposium participants, notably Dan Douglas and Dale Lange, Nina Garrett stressed the priority of theory construction and testing in determining the role of linguistic features in the acquisition of functional skills. Douglas noted that, because of the complexity of verbal communication, it cannot be claimed that the mastery of any set of linguistic features corresponds to a given degree of communicative effectiveness.

Lange pointed out that a glaring flaw in the Guidelines is that their experiential base consists of observations of adult learners. Extrapolating from this base to classroom learning involving children, adolescents, and young adults entails the assumption that FL learning processes remain con-

stant across these various groups. I might add that another weakness involves extrapolation from the specific-purpose instruction characteristic of the various government agencies concerned with FL training and the fundamental instruction that forms the bulk of school and university FL teaching.

John de Jong, representing the Dutch agency for educational testing (CITO), suggested that at lower levels of proficiency, learners are more dependent on structural features and that, consequently, at these levels test might need to be criterion-referenced and discrete-point. More advanced learners have acquired sufficient strategic competence to compensate for structural deficiencies, and their proficiency can be assessed on a single dimension. These facts would lead us to question whether an instrument originally devised by FSI for evaluating a relatively high level of proficiency (S-3 on the FSI scale and Superior on the ACTFL scale) should be the model for assessing the much lower levels of competence realistically attainable by academic learners.

Another moot issue in the evaluation of FL proficiency is the construct of educated native speaker (ENS) speech, which constitutes the implicit target norm in the ACTFL Guidelines and the explicit one for the ILR version. In my own position paper, I argued that in classroom FL instruction, such a target is too restrictive, and I provided as an alternative a multitarget model of language variation according to which native speakers differ with respect to a variety of norms and the degree to which they are influenced by shared norms in particular social contexts. According to that model the speech of native speakers also varies with regard to relative planning in response to another set of social and contextual factors. I suggested that the special social context represented by the FL classroom warrants the establishment of special pedagogical norms. Because it involves the presentation in pedagogical materials of sociolinguistically stigmatized features, the notion of pedagogical norm raises the thorny issue of accuracy and fossilization. Sociolinguistic deviance is unavoidable in early-level interlanguage; even speech targeted on ENSs may be deviant in certain social contexts. The issue of the terminal 2+, around which discussions about the importance of grammatical accuracy have turned, will generate more heat than light until a solid body of empirical observations can be adduced in support of the claim that allowing learners to produce communicatively adequate but structurally deviant utterances inevitably leads to fossilization. Perhaps fossilization can be reduced by a better match between communicative demands and linguistic means available to learners. But instead of reducing functional demands in the interest of more accurate grammatical expression I suggested that, whenever possible, the level of grammatical complexity be adjusted to the learners' limited competence.

4. Administrative and practical problems

The implementation of a national metric for the assessment of functional ability in FLs poses numerous administrative and budgetary problems. Few institutions of higher learning, and even fewer school systems, are equipped to administer a resource-intensive test of functional ability. What is the most efficient administrative structure for training qualified testers and validating training of individual teachers? Of course, the answers will vary depending on the number of learners of a given language. Whereas there are ten of thousands of learners of French, German and Spanish, only a few dozen persons each year attempt to acquire a functional ability in such languages as Dutch, Haitian Creole, or Indonesian.

Another set of questions that must be answered concerns the various uses of a national metric. At what levels of study (beginning, intermediate, advanced) should learners be administered resource-intensive tests and for what skills? Should a national metric be used for validation of foreign language study at the secondary school level? For placement in university courses? For university language requirements? For certification of language majors or prospective secondary school teachers? For admission to specialized graduate school programs?

More than five years after its introduction at the high school and university level the OPI is still not widely used. The somewhat trial nature of the test results in large part from the immense logistic and financial problems posed by its widespread and massive use. Nonetheless two noteworthy attempts have been made to assess FL learning by the use of a proficiency test. An important part of the Symposium on the Evaluation of Foreign Language Proficiency were progress reports of the attempt to institute language proficiency requirements at the University of Pennsylvania and at the University of Minnesota.

Barbara Freed described the University of Pennsylvania's pioneering adaptation in 1981 of a variant of the FSI test and ILR guidelines as a proficiency-based exit requirement in fourth-semester FL courses. The initiative required a two-year preparatory period during which administrative procedures were devised, student fears allayed, suitable instruments implemented, and appropriate cut-off levels determined. Although the introduction of proficiency-based exit requirements have not led to a demonstrable improvement in student performance after two years of study, it has invigorated the FL programs and markedly improved articulation between lower-division courses by providing long-term objectives independent of specific course content. Most importantly, it has shown that an evaluation instru-

ment such as the OPI that makes special demands for staff time and training and that induces considerable student anxiety can indeed by integrated into the curriculum of mid-size university programs.

As was the case with the University of Pennsylvania program, the University of Minnesota initiative described by Dale Lange was motivated by an educational policy decision to define achievement in FL study in terms of demonstrable proficiency rather than time spent in study. The Minnesota program is a bolder experiment because it involves a redefinition of both entrance and graduation requirements in proficiency terms. A typical midwestern state university, Minnesota has larger lower-division FL programs than those of private institutions like the University of Pennsylvania. Also, changes in the entrance requirements of public universities necessarily have a backwash effect on local high school graduation requirements, and they require close coordination between administrators and teachers at both levels. These differences are reflected in the form of the test itself and also in the guidelines that underlie the proficiency ratings and the levels of minimal performance. As was the case for Barbara Freed, Dale Lange did not adopt the ACTFL Guidelines unquestioningly. On the contrary, he underscored some of their shortcomings, particularly for the reading requirement, and he recommended continued validation and revision. He concluded with a thorough and candid analysis of the problems that the implementation of any integrated proficiency-based FL program faces given the many inadequacies of prevalent teacher training, curriculum and syllabus design, materials development, and, most importantly, administrative practices in major university FL programs.

The exit proficiency levels set at the University of Pennsylvania and at the University of Minnesota lead one to question the emphasis that the Proficiency Movement places on developing a national metric for the assessment of oral interaction skills for what I have termed basic, fundamental FL instruction. At both institutions, the oral proficiency level is fixed at the Intermediate Mid level which is defined in the old Generic Guidelines as the ability to satisfy most survival needs and limited social demands in a manner that is intelligible to native speakers used to dealing with foreigners (the revised 1986 Guidelines are couched in somewhat different terms: the ability to handle successfully a variety of uncomplicated, basic and communicative tasks and social situations and to be understood by sympathetic interlocutors). At the University of Minnesota entrance oral proficiency levels are set at the Novice High level which is defined in the 1986 Generic Guidelines as the ability to satisfy partially the requirements of basic communicative exchanges by relying heavily on learned utterances but occasionally expanding

these through simple recombinations of their elements. Learners at the Novice High level will have difficulty being understood even by sympathetic interlocutors. It is legitimate to ask what educational purpose is served by a test designed to measure levels of attainment so low as to pose acute problems of reliability and, ultimately, of fairness.

The gain in oral interactive proficiency between entrance and exit that the University of Minnesota testing program attempts to measure is relatively meager. Such a testing program could meet the criterion of fairness only if it could be demonstrated that the assessment met high standards of reliability. However Barbara Freed pointed out that, although the Pennsylvania proficiency test adheres to the guidelines and format of the OPI, it is administered and scored by teaching assistants who generally have not undergone the rigorous training required for ACTFL certification. Because the University of Minnesota oral proficiency tests need to be administered to hundreds of students, it takes the form of the ROPE (Recorded Oral Proficiency Examination) rather than the standard OPI. Although it is evaluated in terms of the ILR guidelines, the ROPE (Lowe & Clifford) uses a stimulus audiotape on which conversational questions of increasing complexity and sophistication are asked by a native speaker. As is the case at the University of Pennsylvania, the scoring is handled by relatively untrained graduate teaching assistants rather than by certified teachers.

Protase Woodford emphasized the need for compromise in the generalization of a common metric. He pointed out that the University of Pennsylvania's pioneering initiative required setting standards below those needed for students to reach the functional objectives selected. For instance, attainment of the Advanced level in reading would scarcely be sufficient for the analysis of a piece of literature. He suggested the elaboration of two sets of scales. One set would provide teachers with detailed information for diagnostic purposes; a second set, couched in more general performance objectives, would be addressed to users.

John Clark and Ray Clifford mentioned attempts by government agencies to develop variants of the OPI that are less onerous to administer and that measure linguistic skills other than interactional ability.

Conclusion

Without denigrating the considerable accomplishments resulting from the various initiatives undertaken by ACTFL, it is fair to say that much remains to be done before the FL teaching profession has at its dis-

posal a common metric to measure proficiency in the functional use of language in its various modalities that meets generally recognized canons of face validity, reliability, and fairness and which, at the same time, is relatively easy and economical to administer on a large scale. Even more remains to be done before we can substantiate any claims about direct links between the attainment of various levels of proficiency and stages in second language development, on the one hand, or between specific levels of proficiency and inventories of discrete linguistic, discourse, or sociolinguistic features that need to be mastered, on the other hand.

In his perceptive synthesis of the Symposium G. Richard Tucker remarked that much of the controversy surrounding the adaptation of the ACTFL Guidelines and the OPI to what I have termed the fundamental FL teaching programs of high schools and universities (the teaching of FL as part of a liberal education) is attributable to an attempt to meet a variety of incompatible objectives and to please a disparate clientele. He felt that progress in the development of more valid and reliable testing procedures required collaborative work in three areas. First, operational definitions of basic terms such as "fluency", "accuracy", and even "proficiency" need to be developed. Indeed, Pardee Lowe considered this to be the most important challenge that the FL teaching profession must take on. Without precision and common agreement in the use of basic terms, exchange of information and constructive discusson are nigh impossible. Second, a proactive research program for the validation of the OPI and the Guidelines patterned on that which has accompanied the development and the refinement of the TOEFL test needs to be instituted. I have indicated some suggested topics for a vigorous program of research and development earlier in this report, but let me mention several others: (1) comparing the scores of various proficiency tests with actual performance in real world language use tasks; (2) submitting proficiency test scores to secondary analysis, for example, having academic testers score tests rated by government agencies raters and vice-versa; (3) determining to what extent the ability to make reliable OPI rating remains constant over time and exmining the relative resistance of various types of linguistic features to attrition and language loss; and (4) exploring which linguistic features are more likely to trigger stereotypic reactions on the part of native speakers or those who use test results. Finally, Tucker made a plea for wider sharing of information, not only within this country but on an international basis.

At the symposium on the Evaluation of Foreign Language Proficiency, a number of concerns about the direction taken by the development of the OPI surfaced. The Proficiency Movement is still in its incipient stage.

Given many of the gaps in our understanding of the nature of FL proficiency and of processes of FL learning, it seems hazardous to continue to focus on stressing the evaluation of communicative interaction. It is also risky to invest in developing the OPI to measure low-levels of proficiency. The amount of interactional skills that can be developed in our short high school language sequences or in elementary and intermediate university courses is too small and too evanescent. It would be naive to assume that when they discover that the typical four year high school FL program will only develop the ability to meet survival needs and limited social demands, students and administrators will clamor for longer periods of study that can guarantee meaningful communicative ability. On the contrary, such a realization may erase current positive attitudes toward FL study or lead to a backlash against emphasis on oral skills. It would be better to invest in developing proficiency tests for those skills, such as listening comprehension and reading, where more significant results can be shown after modest periods of study. Also, these skills are more robust and less likely to be eroded than communicative ability. Speaking proficiency could be evaluated by the various types of semi-direct procedures described by John Clark and Ray Clifford, and the evaluation should serve mainly diagnostic purposes.

More importantly, the limited means at our disposal should be marshalled for the development of valid and reliable tests for the evaluation of students who are learning FL's for specific purposes: undergraduate langauge majors, graduate students, and prospective teachers. Early indications of foreign language education majors revealed that nearly half of those tested ranged from Intermediate High to Novice High on the ACTFL scale or 1 + to 0 + on the FSI scale. Alas, we must assume that this shocking state of affairs may not have improved appreciably. Our first priority must be the development of proficiency tests which specialists in testing and second language learning believe to be sufficiently valid, reliable, and fair to be made an integral part of the training of our FL specialists, including those entrusted with the teaching of FL's as a fundamental part of a general education.

Notes

[1] The following positions papers were presented (discussant names are listed after each paper): John L. D. Clark, Ray T. Clifford, "The FSI/ILR/ACTFL Proficiency Scales and Testing Techniques: Development, Current Status, and Needed Research;" Judith E. Liskin-Gasparro, "The

ACTFL Proficiency Guidelines: An Update;" Lyle F. Bachman, "Problems in Examining the Validity of the ACTFL Oral Proficiency Interview," (Sally Magnan, Elana Shohamy); David Hiple, "The Extension of Language Proficiency Guidelines and Oral Proficiency Testing to the Less Commonly Taught Languages," (Ron Walton, Sally S. Magnan, Patrick Bennett, Ernest N. McCarus, Jos Nivette, David Singleton, Irene Thompson, John U. Wolff); Dan Douglas, "Testing Listening Comprehension in the Context of the ACTFL Proficiency Guidelines," (John H. A. L. de John, Harry L. Gradman); Albert Valdman, "The Problem of the Target Model in Proficiency-Oriented Foreign Language Instruction," (Pardee Lowe, Jr., John U. Wolff); Heidi Byrnes, "Features of Pragmatic and Sociolinguistic Competence in the Oral Proficiency Interview," (Dell Hymes, Claire Kramsch); Nina Garrett, "The Role of Grammar in the Development of Communicative Ability," (Simon Belasco, Theodore V. Higgs); Susan M. Gass, "L2 Vocabulary Acquisition," (James R. Child, Sharon Lapkin, Susanne Carroll); Barbara F. Freed, "Issues in Establishing and Maintaining a Language Proficiency Requirement;" Dale L. Lange, "Developing and Implementing Proficiency Oriented Tests for a New Language Requirement at the University of Minnesota: Issues and Problems for Implementing the ACTFL/ETS/ILR Proficiency Guidelines," (Robert C. Lafayette, Sandra Savignon, Protase Woodford); G. Richard Tucker, "Evaluation of Foreign Language Proficiency: Synthesis."

[2] Albert Valdman, ed. *Proceedings of the Symposium on the Evaluation of Foreign Language Proficiency*. Bloomington, Ind.: Committee for Research and Development in Language, Indiana University, 1987.

Bibliography

Freed, Barbara F. "Issues in establishing and maintaining a language proficiency requirement." *Proceedings of the Symposium on the Evaluation of Foreign Language Proficiency*. Ed. Albert Valdman. Bloomington, IN: CREDLI:263-274.

Kramsch, Claire. "From language proficiency to interactional competence." *Modern Language Journal* 70(1986): 366–72.

Lowe, Pardee Jr. & Ray T. Clifford. "Developing an indirect measure of overall oral proficiency." *Measuring Spoken Language Proficiency*. Ed. James R. Frith. Washington, DC.: Georgetown University Press, 1980: 31-39.

Omaggio, Alice A. *Teaching Language in Context*. Boston: Heinle and Heinle, 1986.